

Stereo Recording & Rendering – 101

http://www.linkwitzlab.com/Recording/AS_creation.htm

- 1.0 Binaural - Directly from eardrum to eardrum - 1
 - 1.1 Binaural in practice - 2
 - 1.2 Quasi-binaural recording for loudspeaker playback - 2
- 2.0 From mono to stereo by panning - 4
 - 2.1 Recording and rendering with paired microphones - 6
 - 2.2 Rendering below 1 kHz - 10
 - 2.3 Spaced microphones - 13
 - 2.4 Blending AB and XY recording techniques - 15
- 3.0 Rendering in a reverberant environment - 16
 - 3.1 Loudspeaker placement - 20
- 4.0 Summary - 21
- 5.0 References - 21

1.0 Binaural - Directly from eardrum to eardrum

In its most accurate, but highly impractical form, a binaural recording captures the eardrum signals of a particular person at an acoustic event. The signals are then played back for the same person using headphones, which have been equalized so that the identical signal is reproduced at the eardrums as was present there during recording, Figure 1. The same small microphones (behind thin, flexible plastic tubes almost touching the eardrums) as used during recording, are also used to equalize the headphones for playback. It has been reported that on playback the aural scene is localized outside of the head, even when in frontal line of sight [1]. If the head is turned sideways during recording, then the aural scene turns sideways during playback. But if the head is turned during playback, then the eardrum signals do not change and the position of the aural scene moves with the head, which is unnatural. This tends to collapse the aural scene to a location inside the head.



FIG. 1 Binaural recording and reproduction of the eardrum signal

Note that the frequency responses of head, pinna and ear canal of the listening individual are imbedded in the recording. Equalization of the headphones removes enclosed pinna and ear canal resonances in order to reproduce the recorded signal exactly at the ear drum, but only for that one individual.

1.1 Binaural in practice

In practice an artificial head is used, with outer ears of generic shape and texture. Neck, shoulders and upper body are included for scientific work' Figure 2. Microphones are usually mounted at the ear canal entrance. The microphone signal frequency response describes the Head-Related-Transfer-Function, HRTF, for blocked ear canals of this mannikin.

The microphone signals may be reproduced via over-the-ear, on-top-of-the-ear or in-the-ear headphones and thus be rendered differently depending upon the type and make of the phones. Localization of the aural scene is typically inside the head or very near the head. The aural scene stays locked to the head as the head is moved or turned. This is unnatural and the major flaw of binaural. It is completely overcome with a head tracking system. There is also a lack of tactile perception of vibration even when the sound volume is very high. But binaural can provide a very high degree of realism due to low non-linear distortion and being isolated from ambient sounds. You can have the sensation of someone whispering into your ears, which is impossible to achieve with stereo loudspeakers.



FIG. 2 Artificial head, opened to show microphone mounting fixture and outer ear

1.2 Quasi-binaural recording for loudspeaker playback

It would seem natural that binaural recordings should work well for stereo loudspeaker sound reproduction. But the frequency response above about 3 kHz of the microphone pickup at the blocked ear canal is strongly determined by the shape and detail of the outer ear and the direction of sound incidence. This is the Interaural-Level-Difference, ILD frequency range of directional hearing. When this signal is reproduced over loudspeakers the response variations impart a coloration and misleading directional cues to the signals arriving from $\pm 30^\circ$ at the listener's ears. Stereo recording microphones therefore have no pinna. The shape of the head and the distance between the ears causes sounds to arrive at the ears at slightly different times depending upon the direction of the sound source. Arrival times differ by less than 700 ms, which is used to give the brain directional cues in the Interaural-Time-Difference, ITD frequency range of hearing below 800 Hz. ITD and ILD are elements of the Head-Related-Transfer-Function, HRTF.



FIG. 3 Microphone capsules placed on the surface of a 19 cm sphere

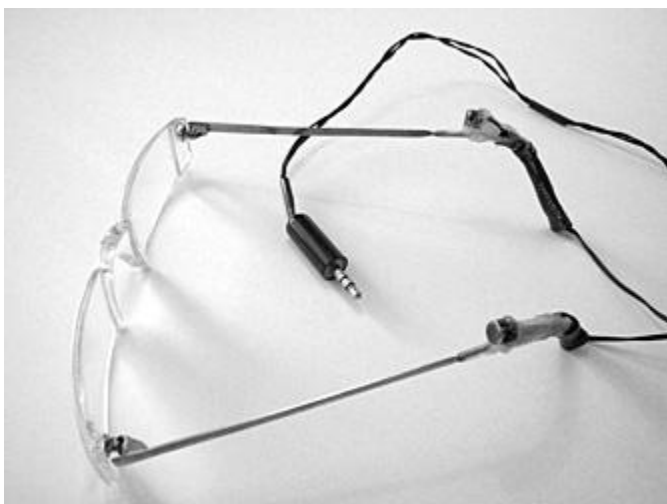


FIG. 4 Omni-directional microphone capsules placed close to the ears, but not inside the pinna

A sphere microphone preserves the spacing between the ears for low frequency sound localization and provides blockage of sound between the ears for the higher frequencies, Figure 3, [2]. I have used this microphone and a similar setup for my own head, Figure 4, to make recordings of events where I also then had a direct memory of what I heard. This gave me material to evaluate my loudspeaker designs.

Sphere microphone are not optimal for rendering the recorded aural scene over two loudspeakers. The spacing between the microphones and the lack of low frequency directivity introduces misleading spatial cues in the ITD range of hearing and affects the realistic creation of phantom sources and aural scene. This will be shown below where loudspeaker stereo is explained.

2.0 From mono to stereo by panning

Stereo over two loudspeakers works by creating a phantom acoustic scene between the loudspeakers. In its simplest form a single monaural signal is fed to both left and right loudspeakers. If the two loudspeaker levels are identical and there is no phase difference between them, then a listener on the center line between the loudspeakers will hear the monaural sound as coming from the center though there is no sound coming from that direction. This is basically a very unnatural event and so a slight movement to the left or right will shift the phantom source to the nearer loudspeaker. If the monaural signal is artificial, like pink noise, then there will also be a change in tonal color at certain off-center locations due to interference of left and right loudspeaker signals at the ears. This has been called "the fundamental flaw of 2-channel stereo" and a reason for promoting a center channel in 5.1 surround [3]. But surround has its own phantom source issues, between L and C, between C and R and more.

While the off-center interference is audible with pink noise, it is not an issue with signals that are known to the brain like music or voice and when there is a reverberant sound field due to the listening room. The timbre of sound from a center loudspeaker can be somewhat different from the timbre of the corresponding center phantom source, because sound arrives from different angles at the ears in the two cases. Thus their HRTF and the number of sound streams is different, but the brain compensates effectively.

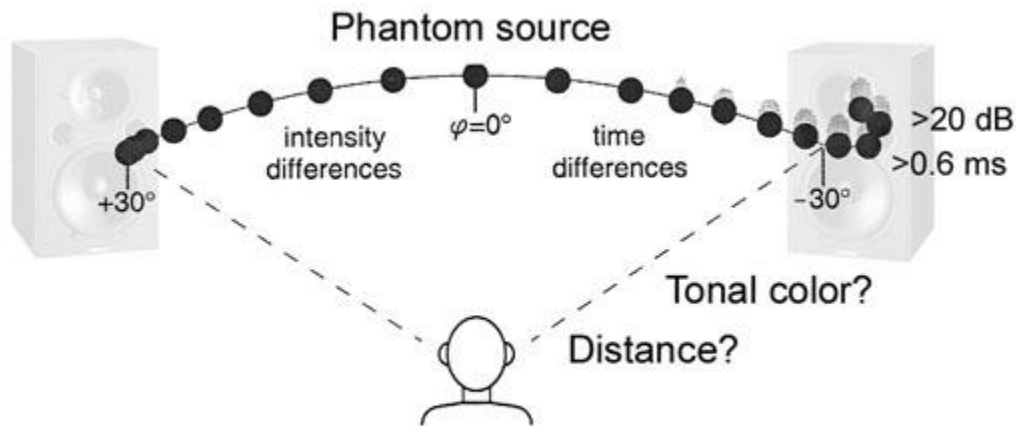


Fig. 1.4. Perceived directions with pink noise, constant loudness

Fig. 5 By reproducing a single channel signal over two loudspeakers a phantom source is perceived between the loudspeakers [4]

Primarily level panning is used for recording because it provides a more defined phantom source, Figure 5 and Figure 7. Today the majority of recordings are produced by down-mixing a multiplicity of monaural tracks of different instrument pick-ups into a 2-channel or n-channel format, Figure 6. Today, the large mixing board is implemented in digital form on a computer with a large screen display.

The mix-down often involves equalization, the addition of reverberation and compression in order to fit the taste of the recording engineer and market expectations as seen by the producer.

Panning distributes phantom sources along a line between the loudspeakers. The distance of the phantom scene from a listener is essentially given by the distance between listener and loudspeaker. If the loudspeakers are highly directional and the room is acoustically dead, then the image is sometimes closer than the loudspeakers, approaching headphone listening. Depth and

height behind the loudspeaker line depends upon cues that the brain receives from reverberation and volume levels of sources in the recording. It is difficult to produce a spatially coherent mix from multiple tracks that is believable. The result is typically a collage of phantom sound clusters next to and on top of each other, or a wash of diffuse sound.



Fig. 6 32-channel mixing-board and recording monitors to create a phantom acoustic scene from up to 32 individual signals

Since the majority of loudspeakers interact with the listening room unfavorably and they are also set up too close to the walls, few listeners notice the spatial distortion in the recording. They probably have never experienced the realism that 2-channel stereo is capable of when recording and reproduction are executed optimally.

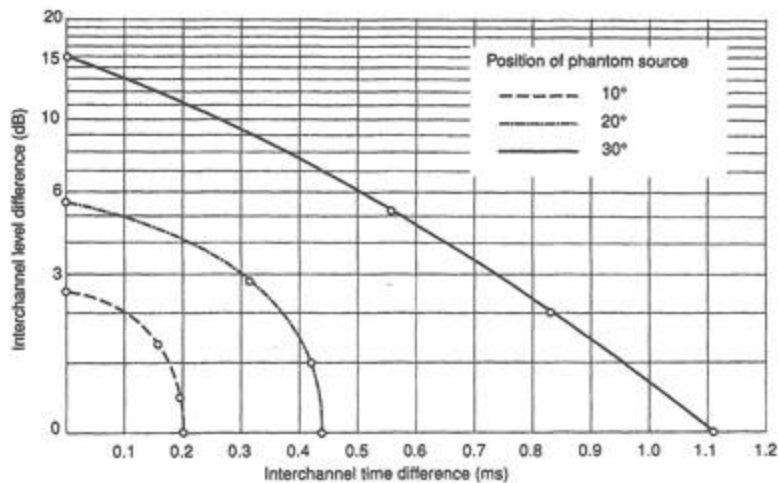


Fig. 7 Placement of the phantom source between the loudspeakers depends upon level and timing differences between the loudspeaker signals [5]

2.1 Recording and rendering with paired microphones

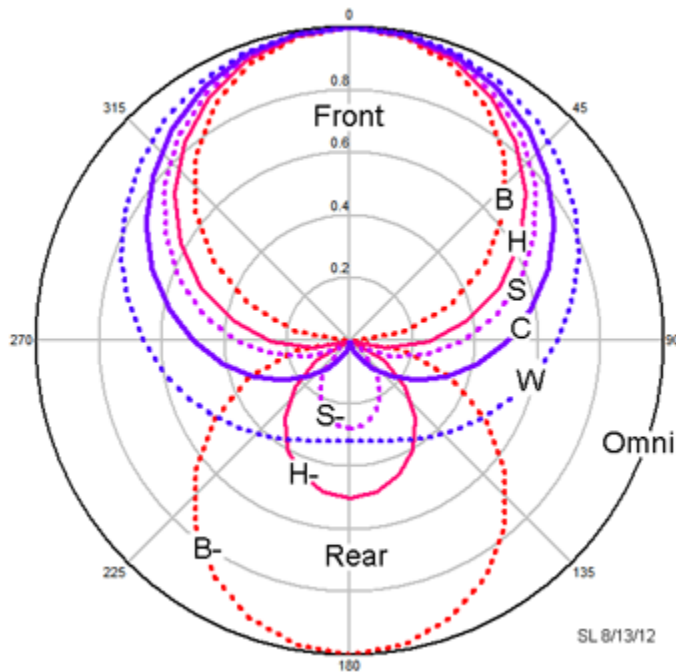


Fig. 8 Polar responses of Omni-directional, Wide-cardioid, Cardioid, Super-cardioid, Hyper-cardioid and Bi-directional microphones with identical outputs at 0° sound incidence angle. Output polarity is reversed with three of the microphone types for sound incidence from the rear.



Fig. 9 Polar responses of two cardioid microphones at $\pm 55^\circ$ angle to each other in the horizontal plane and mounted on top of each other for coincident sound arrival at the

microphones. A signal at 25° incidence, for example, produces a larger output level from the right microphone than from the left microphone.

Microphones are built with a wide variety of directional characteristics. A few types are shown in Figure 8. The directional behavior is usually frequency dependent to some degree and more so with large diameter and large area microphones. An issue with directional microphones is decreased sensitivity to lower frequencies, high sensitivity to wind noise, pop and shock. Unlike omni-directional microphones, which are sealed sound pressure sensors, directional microphones result from a combination of omni- and bi-directional elements. The bi-directional element is sensitive to sound particle velocity and its direction.

A coincident pair of directional microphones pointing in different directions will have outputs that are in-phase but differ in amplitude when sound is incident at angles other than 0° or 180° , Figure 9.

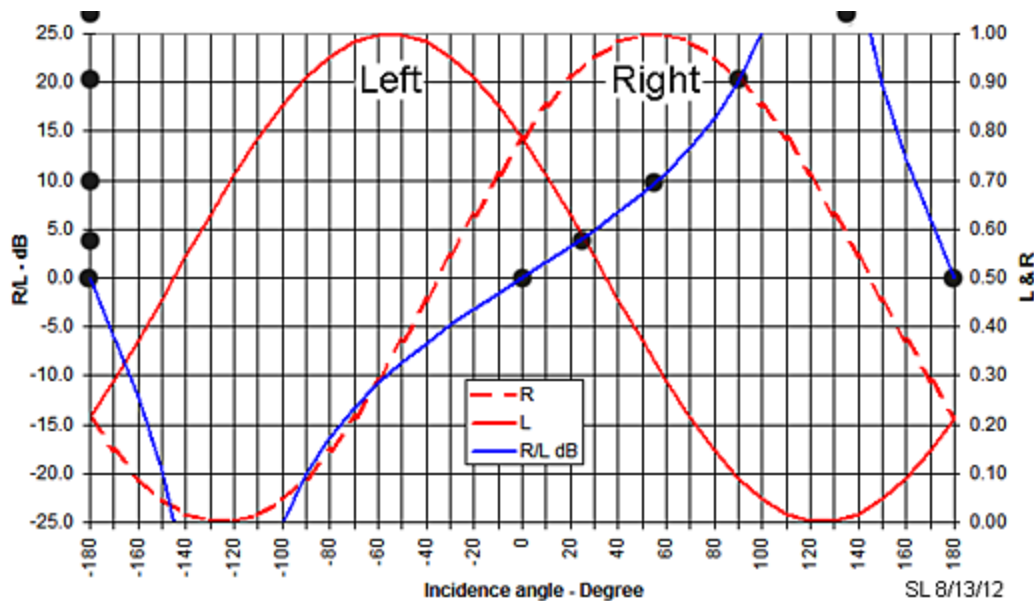


Fig. 10 The polar diagram of Figure 9 has been redrawn with a linear axis for the angle of sound incidence and the ratio of right to left outputs in dB has been added. The ratio determines the position of a corresponding phantom source between the loudspeakers according to the panning law of Figure 7.

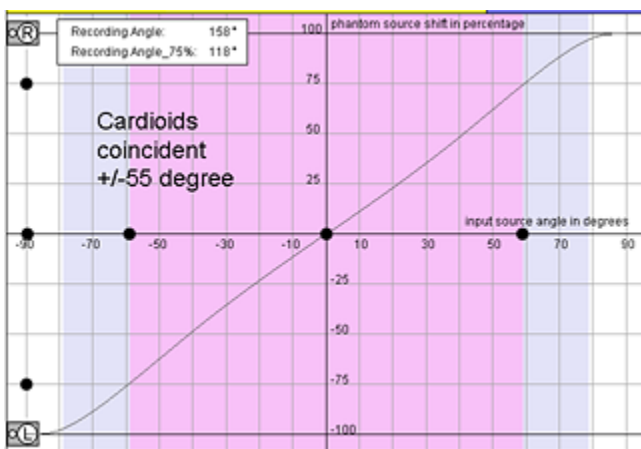
For example, at 25° incidence we see that $R/L = 4$ dB. Using this value and 0 ms time difference between the channels in Figure 7, we find that the phantom source should be about 15° to the right of center or 50% towards the right loudspeaker at 30° . The 10 dB level difference from a signal at 55° will produce a phantom source at about 25° or 83%. The 90° and 135° signals map into the right loudspeaker. Any signal at 180° will show up as a center phantom.

Any pair of angled, coincident, directional microphones will automatically pan the recorded acoustic scene onto the line between left and right loudspeakers as a phantom scene. It will pan into left and right loudspeakers as monaural signals those portions of the $\pm 180^\circ$ view, which do not translate into phantom sources, because the microphone output levels differ too much. The monaural signals are problematic because they set hard boundaries to the phantom scene, which usually is not in keeping with the spread of the aural scene. It draws attention to the loudspeakers and their location, rather than letting the loudspeakers disappear from perception with eyes closed.



Fig. 11 Near-coincident pair of microphones. They are mounted at an angle to each other and to a stand with two shock mounts.

Different types of directional microphones arranged as XY coincident, as near coincident or spaced pairs will lead to a different distribution of the recorded acoustic scene on a standard $\pm 30^\circ$ loudspeaker setup for reproduction, Figure 11. It is important to know the "Stereo Recording Angle" of a particular pair of microphones, because it determines which portion of the 360° acoustic scene to be recorded will show up between the loudspeakers and with what linearity of distribution between the $\pm 75\%$ points symmetrical to the center location. This has been calculated and catalogued with a JAVA applet following a panning law like in Figure 7, [6]. Results for a pair of cardioid microphones are shown in Figure 12.



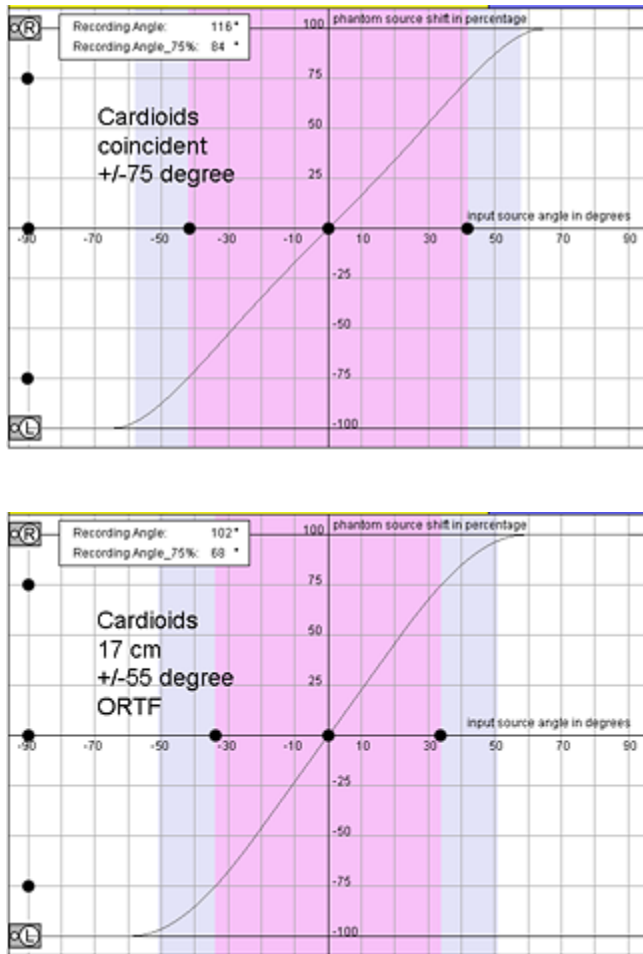


Fig. 12 Stereo recording angle for three different arrangements of cardioid microphone pairs.

The phantom positions for the $\pm 55^\circ$ coincident pair must have been derived from a different panning relationship than in Figure 7. At 25° incidence the phantom source is at 30% or 9° off center whereas from Figure 10 and Figure 7 it would be at 50% and 15° . It should be noted that panning relationships have been determined empirically. Different investigators report different results, also depending upon source material. Directional hearing is a brain process and subjective. Nevertheless the graphs can show trends. Widening the angle to $\pm 75^\circ$ places a narrower $\pm 42^\circ$ section of the acoustic scene between the 75% points, whereas before it were $\pm 59^\circ$. Separating the microphones by 17 cm leads to the well know ORTF arrangement. It further narrows the 75% rendered angle to $\pm 34^\circ$. The graphs only cover the $\pm 90^\circ$ frontal plane, not $\pm 180^\circ$ and elevation. It also cannot be deduced from them how recorded and monaurally rendered left and right loudspeaker signals might affect spatial perception. What exactly then are the aural differences between the three arrangements in Figure 12?

The SCHOEPS Mikrofone website provides sound samples of recordings, which use the same recording angle, but where the types of microphone pairs differ [7]. In general, only an experienced recording engineer with adequate monitor loudspeakers might know how direct, reflected and reverberant sounds from various directions combine in a particular venue at the microphones and will be rendered over two loudspeakers. The monitor loudspeakers used, their setup and environment influence the decisions made by the recording engineer and limit the timbral, spatial and musical quality of the end product, if they impose their own signature or are not fully trustworthy for the task.

2.2 Rendering below 1 kHz

How does level panning, the sound of the same signal but at different levels from left and right loudspeakers, produce a phantom source between the loudspeakers? I call it a magic trick, because there is no sound-wave coming from the direction of the phantom source. The phantom source is a construct of the brain and based upon the signals at the eardrums. Below 1 kHz, where the wavelengths of sounds become large compared to the size of the human head, the ear signals differ very little in magnitude. But if the sound arrives from directions other than the vertical plane of head symmetry, then the ear signals contain a time difference. It is called interaural time difference or ITD and is a very strong contributor to directional hearing [8, 9]. Interaural level differences, ILD predominate above above 3 kHz where the head and outer ear dimensions become large compared to the wavelengths of sounds.

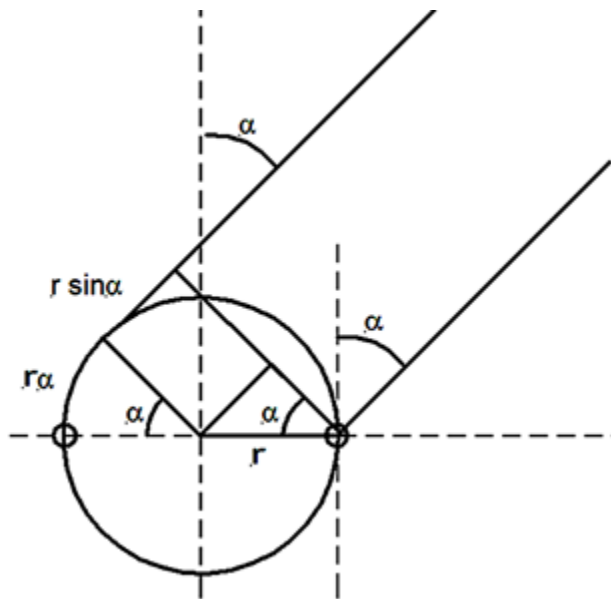


Fig. 13 Sphere model of a head and sound path length from a distant source to left and right ear points

A sphere model of the head without sound diffraction allows for easy calculation of ITD versus sound incidence angle α , Figure 13.

The maximum value of ITD becomes $263 \mu\text{s}$ for a stereo system with loudspeakers at $\pm 30^\circ$, Figure 14.

Each stereo loudspeaker sends its output to both left and right ears where they add as vectors, Figure 15. The phantom source direction, distance, size and sound are derived from the summed sound streams l and r . The frequency response for l and r shows that both ear signals have identical (!) magnitude, Figure 16a. Their level is maximum when L and R loudspeakers have the same output. The level decreases at both ears equally when one of the loudspeakers becomes louder and the other decreases while keeping $L+R = \text{constant}$ [10].

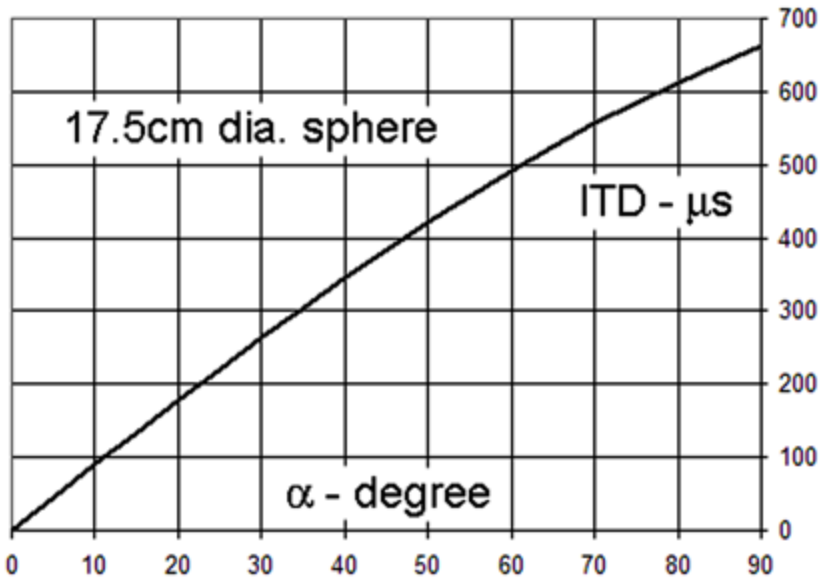


Fig. 14 Interaural time difference as function of sound incidence angle α , where $\text{ITD} = r/c (\alpha + \sin \alpha)$

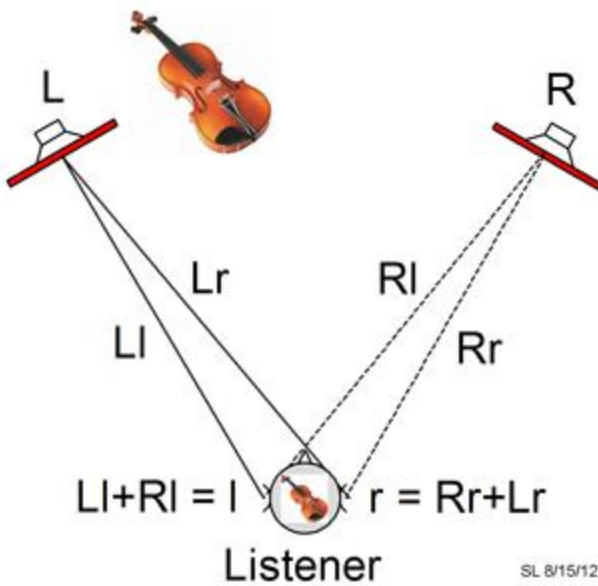


Fig. 15 Loudspeaker signal summation at the ears, which leads to phantom source perception.

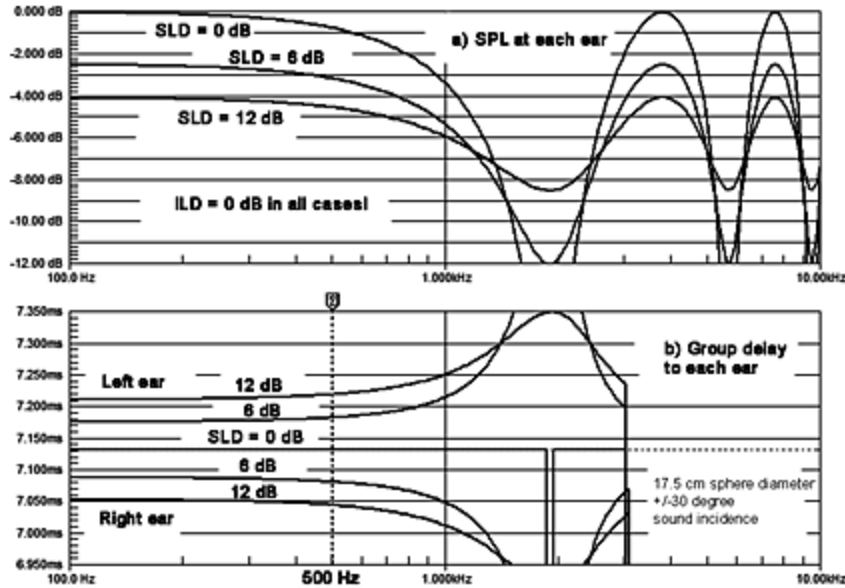


Fig. 16 Frequency response of the summed ear signals. SPL at each ear (a) and group delay (b) as function of source level difference in dB.

The ear signals differ in time of arrival, Figure 16b. ITD is directly related to SLD. Additional values have been calculated for a 500 Hz signal and are plotted in Figure 17 as a function of level differences between the loudspeakers. The ITD values are compared to those in Figure 14 for a real source. The comparison yields the phantom source angle γ , which is thereby known as a function of SLD, [10]. Knowing the output signals and their ratios for a coincident pair of cardioid microphones, Figure 10, we can now derive the phantom source angle γ for a given source angle α and the relative sound level of the phantom source, Figure 18.

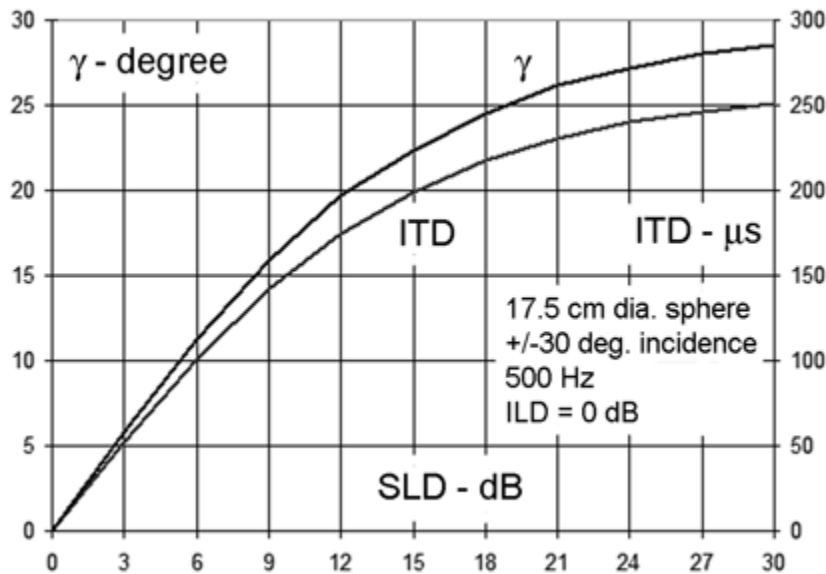


Fig. 17 Interaural time difference and phantom source angle γ as a function of level differences between two loudspeakers at $\pm 30^\circ$.

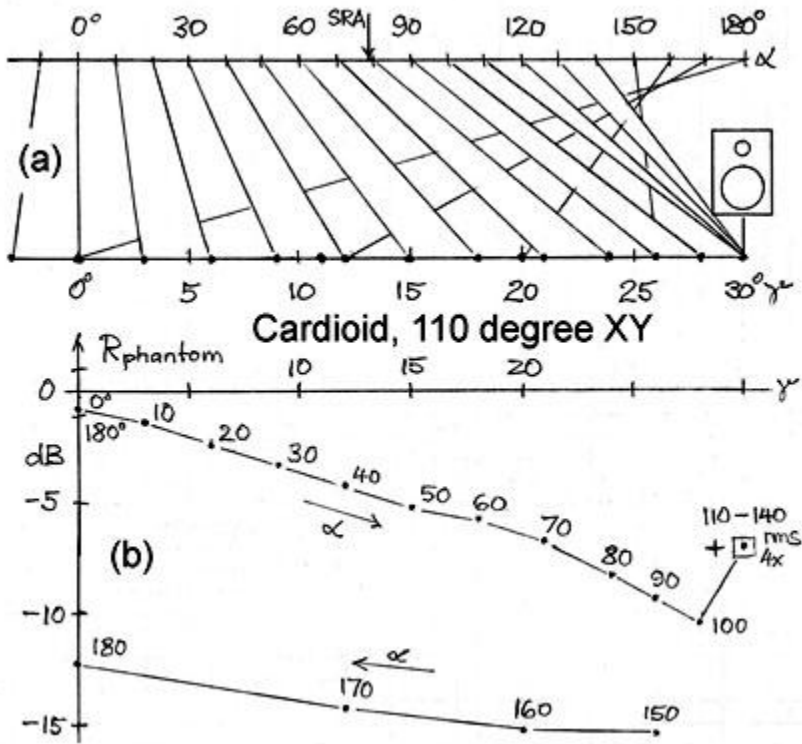


Fig. 18 Stereo rendering of a real source at angle α to a phantom source at angle γ when level panned by a pair of coincident microphones at 110° subtended angle. Transform of angle (a) and level (b) for a constant amplitude real source signal. The -7 dB monaural signal at $\gamma = 30^\circ$ is the rms sum of source signals between 110° and 140° . Linear addition of the four underlying signals would yield a level of -1 dB.

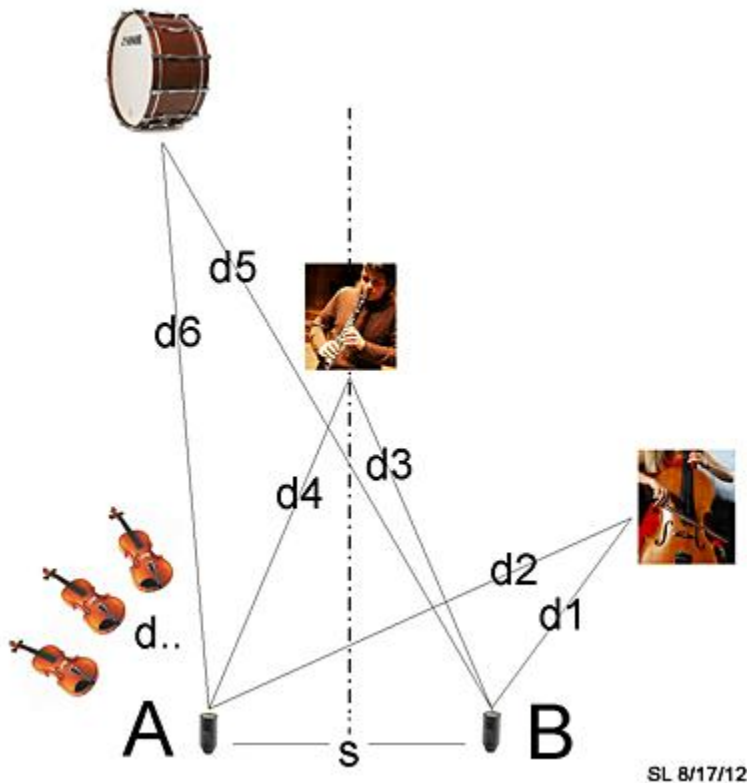
If pure time panning is used, where left and right loudspeaker signals have the same magnitude and only differ in time, then the resulting ear signals have zero (!) time difference. The level at each ear changes in comb filter fashion with frequency and at a different rate for each ear depending upon the time difference. This produces a sense of spaciousness and maybe even a sense of phantom direction, but it is an unnatural phenomenon and peculiar to two loudspeaker stereo. Even near-coincident microphone recordings, as with an ORTF or a sphere microphone setup, would suffer from it to some extent in the ITD frequency range.

In the ILD frequency range the listener's head shadows the $\pm 30^\circ$ loudspeaker signals slightly, which reduces cross-talk and together with the outer ear shape allows for level differences between the ears. Phantom source direction is primarily based on transients in this range.

2.3 Spaced microphones

Spaced microphone recording techniques are prone to large amounts of leakage between sound pickups. In the case of spaced omnis each microphone sees the whole acoustic scene from a different location. Each microphone output emphasizes different sound sources due to greater proximity to one or the other, Figure 19. For example the cello is closer to microphone B than to A and its pickup will be $[20 \cdot \log(d_2/d_1)]$ dB stronger than from A. This ratio will also be modified by the frequency dependent directional radiation properties of the cello [11]. Furthermore the path

length difference $[d_2-d_1]$ causes the output from microphone A to be delayed by $[2.9*(d_2-d_1)]$ ms/m. The sound stream produced by the cello in the output of microphone B can therefore differ significantly in magnitude and phase from that in microphone A.



SL 8/17/12

Fig. 19 Two microphones with with separation s pick up different streams of sound from the same instruments.

The clarinet, which is at equal distance $[d_3=d_4]$ from A and B will produce microphone outputs that still may differ in magnitude due to the directionality of clarinet sound radiation. The path lengths d_5 and d_6 do not differ much for the more distant drum, especially when compared to a wavelength, which is 3.4 m at 100 Hz. But since the drum is a dipolar radiator for many of its membrane vibration modes, it matters how it is facing A versus B. The strength and character of the sound pickup could vary depending upon the separation s between the microphones. The group of violins close to A will be distant in the microphone output B and of a different blend compared to their pickup from A.

The two microphone signals A and B are transmitted to left and right loudspeakers, Figure 20. If we only turn on the left loudspeaker by itself, then we immediately localize the source of the sound as being the left loudspeaker. We are likely to hear that the sound is made up of sound streams from violins, clarinet, drum and cello when we listen for a while. We may develop a sense of how close those instruments were to the microphone and hear the reverberation of the venue when the drum is struck. Similarly when we only listen to the right loudspeaker we will hear the cello more clearly and distinct from the more distant violins. The clarinet may sound just the same. But what happens when both loudspeakers are turned on? Now we experience a phantom scene between loudspeakers and with spaciousness added. Localization becomes diffuse and when many more instruments are involved, tends to cluster around left and right loudspeakers with somewhat of a hole in the middle. Reversing the polarity of one of the loudspeakers will usually not change the aural scene. The problem is for the brain to decipher the two sound

streams $[L+R]$ and $[R+L]$ into something meaningful [12]. Spatial hearing is a process that has evolved in humans over eons of time but there is little brain patterning for these types of sound streams. Still, very enjoyable recordings have been made with spaced microphones for 2-channel stereo. Here the recording engineer must find the location for the spaced pair that yields an aesthetically pleasing recording that does justice to the music's composition and the audience's expectations. Detailed spatial rendering is usually of low priority, also because many playback systems lack that capability.

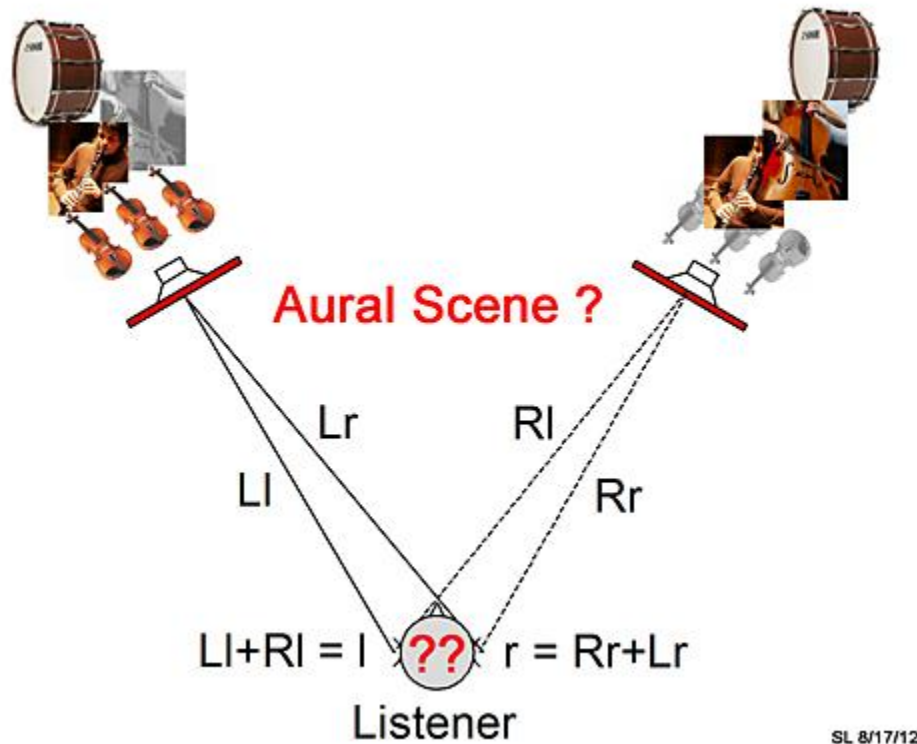


Fig. 20 The summed streams of microphone A are reproduced by the left loudspeaker and those of B by the right loudspeaker. Each ear receives a different sum of loudspeaker signals. The aural scene between the loudspeakers is created in the brain of the listener from the two ear signals.

2.4 Blending AB and XY recording techniques

Instead of using A and B microphones in Figure 19 to produce a 2-channel recording a single microphone could have been assigned to the violins, the drum, clarinet and cello for a 4-channel recording. Unless these microphones are placed very close to their individual instruments they will pick up low levels of all the other instruments and the reverberant sound in the venue. But if a microphone is very too close to an instrument it might pick up a sound that is modified by the directional characteristics of the instrument, that is unnatural and would normally not be heard by an audience [11]. All this can be avoided by placing each musician in an isolation booth and with headphones so he can hear everyone else to play in ensemble. Now we would have four sound tracks that can be level panned between left and right monitor loudspeakers to produce the 2-channel recording. Each track could even be panned multiple times. Violins, drum, clarinet and cello tracks have become sound objects that can be manipulated at will. The result is a collage of sound clusters that hang like sheets of laundry on a clothes line between the loudspeakers. Even when artificial reverberation is added to the mix the aural scene remains spatially flat and unnatural.

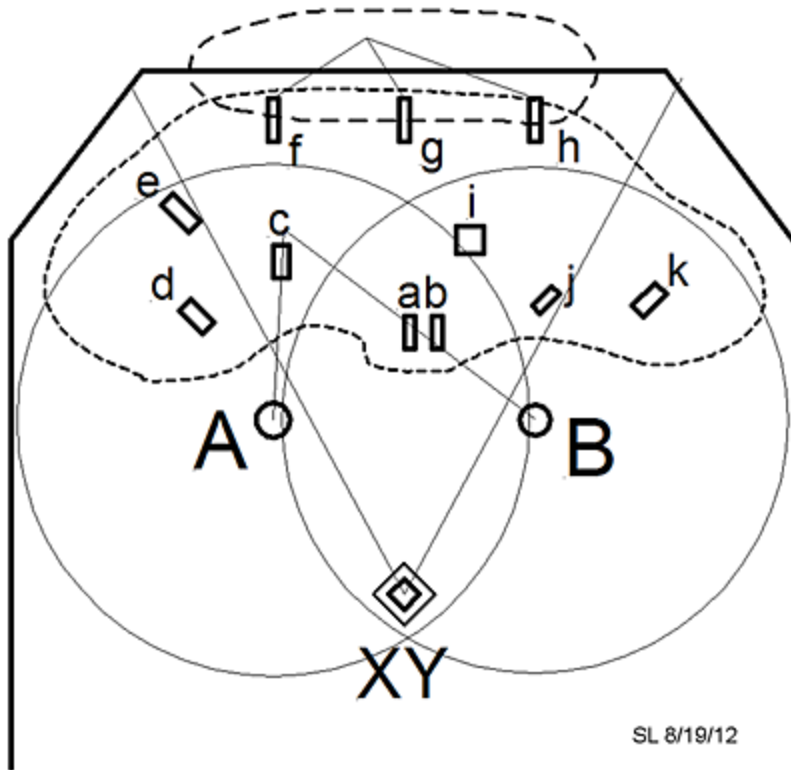


Fig. 21 Microphone setups for preserving natural spatial relationships between sound sources and recording venue

It should be possible to use a single coincident XY microphone pair as the basis for rendering the spatial relationships between individual sound sources and their interaction with the reverberant sound field of the recording venue, Figure 21. The pair must be placed at some distance from the acoustic sources to minimize level differences between near and far instruments and to capture the width and height of the acoustic scene.

If necessary for clarity or audibility, individual sources or groups of instruments (a through k) are recorded as mono signals and then panned to the proper location in the phantom scene established by the XY pair. Electrical signal delay may be needed, if the required amplitude approaches the level of the XY pair output. When more than one microphone (f, g, h) is used to record a larger group of sound sources, then signal overlap and timing differences between the microphone outputs will lead to a loss of clarity when panned to left, center, right respectively. The microphones need to be close to the source to minimize leakage. The panned level of the microphones must be kept low not to override the precedence of the XY outputs. Two microphones (a, b) for a single source will add diffuseness, but only to the high frequency spectrum when placed close together. Widely spaced microphones (A, B) render a diffuse sound depending upon their distance from sources. They are useful for low frequency pickup when further away from the orchestra.

3.0 Rendering in a reverberant environment

Listening to stereo usually takes place in an enclosed space. It means that the loudspeaker signal is reflected by walls and objects in the room not only once but again and again as it bounces around, Figure 22. The signal loses energy with each reflection. Very quickly the sound level in the room, away from the loudspeakers, becomes independent of listener location. That level is

established by the rate at which acoustic power is fed into the room and the rate at which the reverberated sound is dissipated as heat by walls and objects. A listener very close to the loudspeakers will hear primarily the direct sound. At greater distance from the loudspeakers the reverberated and diffuse sound in the room dominates. It is louder than the direct sound at that location. The distance from the loudspeakers, at which direct and reverberant sound levels are equal, is called the reverberation radius or critical distance [13].

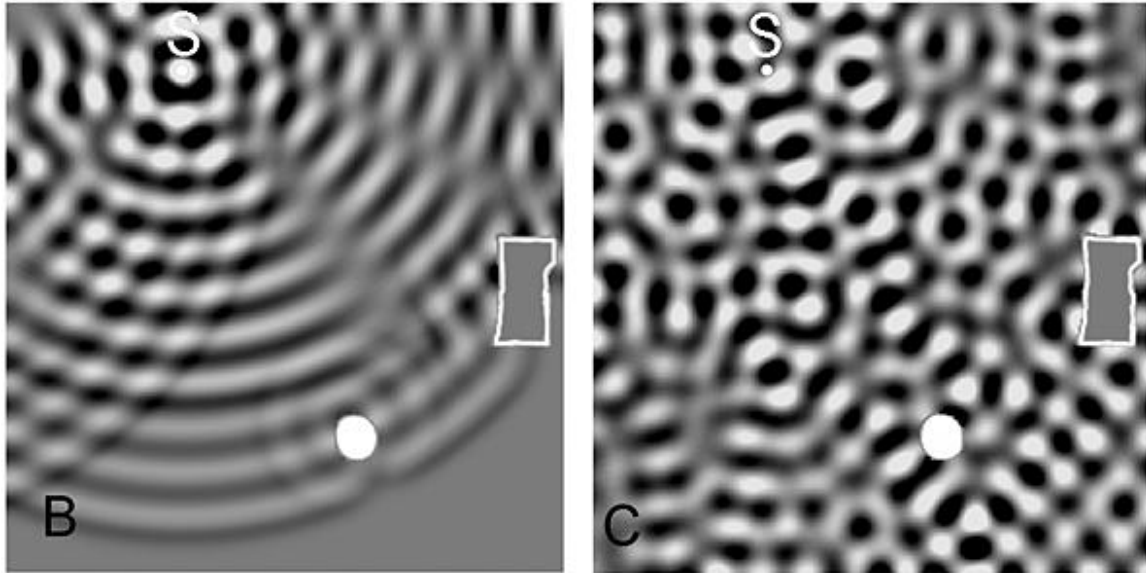


Fig. 22 Wave tank display of a source S radiating a constant wavelength signal into a bounded surface.

(B) Reflection of the initial wave by surrounding walls. Reflection and diffraction by objects. The circular object is in a plane wave or free-field situation for a short time.

(C) Snapshot of the reverberating wave distribution at a later time. The circular object is immersed in a diffuse wave-field.

At low frequencies, where the wavelength of sound is no longer small compared to the dimensions of the room, the reverberant field takes the form of standing waves or room modes. As a consequence, the difference in sound level between two locations in the room can be very large at certain low frequencies. Rendering problems can be minimized by installing acoustic absorbers and/or minimizing the excitation of objectionable modes by woofer placement and/or directivity.

For example, the room in Figure 23 has 21 possible modes below 100 Hz. The degree to which anyone of these is stimulated depends upon loudspeaker placement and radiation pattern. In addition, which of these are heard depends upon the listener's location in the room.

Mode frequencies are widely spaced at low frequencies. They become closer and eventually overlap as frequency increases. At the 120 Hz Schroeder frequency two mode frequencies fall within the 5.7 Hz resonance bandwidth of the modes.

The reverberation time T_{60} is an important measure of a room's acoustic properties. It describes how fast the reverberant sound in the room decays by 60 dB when the loudspeaker is turned off. Knowing T_{60} one can calculate the Schroeder frequency, which is the low frequency boundary of statistical room behavior. One can also calculate the critical distance.

A					
		L =	23.0	ft	7.02 m
Room dimensions in feet		W =	16.4	ft	5.00 m
		H =	9.0	ft	2.75 m
		Floor area A =	378	ft ²	35.1 m ²
		Volume V =	3401	ft ³	96 m ³
		Surface area S =	1465	ft ²	136 m ²
		Volume/Surface V/S =	2.32	ft	0.71 m
		Edge length Le =	194	ft	59 m
		Below frequency fm =	100	Hz	
		Total number of modes N =	21		
		Avg. mode spacing df =	3.4	Hz at fm	
		Estimated reverberation time T60 =	384	ms	
		Resonance bandwidth bw =	5.7	Hz	
		Rise time Trise =	123	ms	
		Schroeder frequency fs =	120	Hz	
		Monopole reverb distance Rm =	0.89	m	35 inch
		Dipole reverb distance Rd =	1.55	m	61 inch
		Avg wall absorption a =	30%		
		Estimated avg wall absorption a =	30%		
		Reverberation time T60 =	384	ms	

Fig. 23 Statistics of a room [13, 17]

A reverberation time of 384 ms means that 30% (41 m² or 440 ft²) of the room surface area behaves acoustically like open windows through which sound escapes. That is more than the floor area in size. A shorter reverberation time requires even more absorption.

The direct SPL from the loudspeaker, which decreases as [1/distance], becomes equal to the reverberant sound level at 0.89 m, provided that the loudspeaker radiates like a perfect omni. The critical radius becomes 1.55 m for an ideal dipole because of its directivity of 4.8 dB.

If we wanted T60 = 250 ms for the above room, then the average wall absorption would have to increase to 54%, Figure 24. This will take considerable effort to achieve and the room will no longer be experienced as a normal living room. A reverberation time of 250 ms is required by EU Broadcasting Standards for recording studios. The benefit of reduced reverberation time is an increase in reverberation radius from 0.89 m to 1.30 m for the monopole and from 1.55 m to 2.25 m for the dipole, Figure 25. Typical studio monitors radiate omni-directional from low frequencies up to several hundredth of Hz. They become increasingly forward directional as frequency increases, when baffle and radiator sizes become first comparable and then large in size relative to the wavelengths of radiated sound. Thus the reverberation radius and the ratio of direct to reverberant sound will increase with frequency compared to the monitor's lower frequency interaction with the studio room.

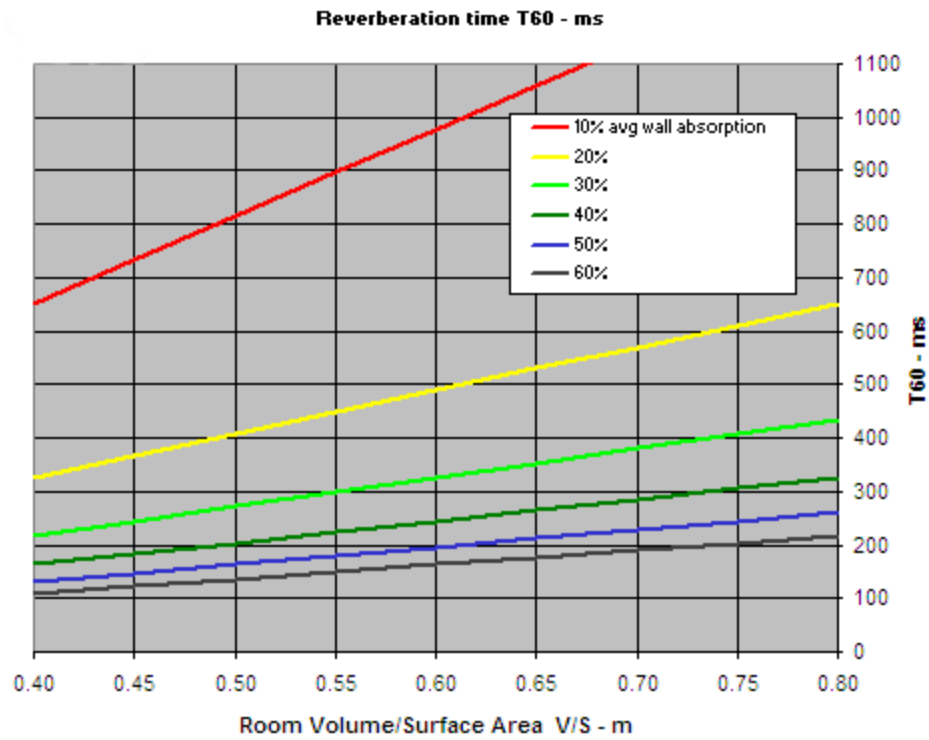


Fig. 24 Reverberation time as function of room volume and surface area for different percentages of wall absorption.

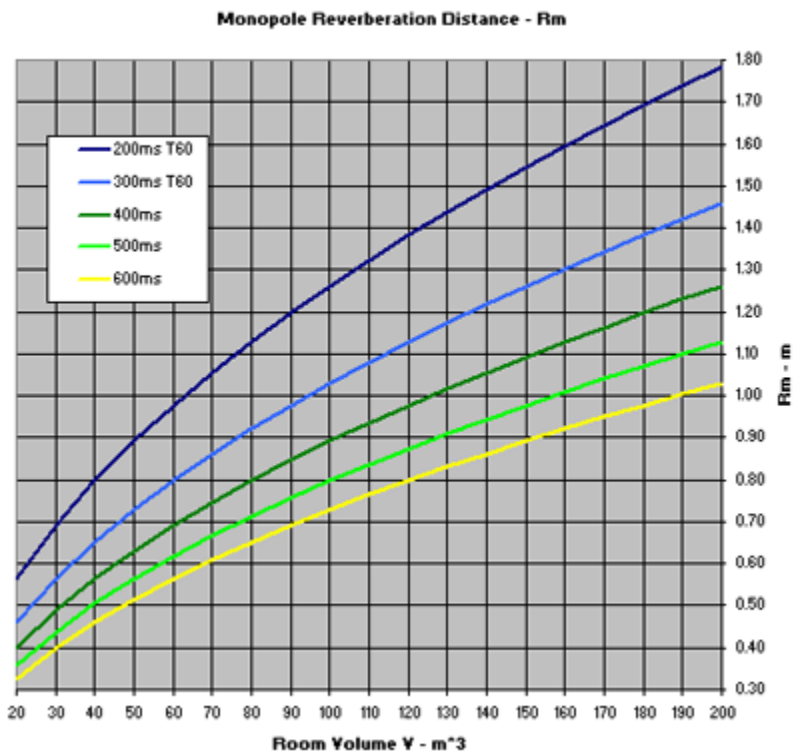


Fig. 25 Reverberation radius as function of room volume for different reverberation times.

A dipole loudspeaker with frequency independent directivity will readily achieve the same direct-to-reverberant ratio in a more lively room as the typical monitor loudspeaker in a recording studio.

Statistical room parameters can be estimated from the spreadsheet modes1.xls [17].

3.1 Loudspeaker placement

The reflective and reverberant environment of a room restricts the placement and radiation characteristics of the loudspeakers. The equilateral triangle, formed by the loudspeakers and the listener in a stereo system setup, must be placed symmetrical to the room boundaries or large objects in the room, Figure 26. This ensures symmetry of the reflections relative to the center axis and a sharply defined center phantom image for monaural loudspeaker signals or center panned sounds. Each loudspeaker also must be placed at some distance from front and side walls so that reflections reach the ears later than the direct loudspeaker sounds. In that case the brain can filter out the direct sound streams more readily from those due to reflection and reverberation. A minimum time gap of 6 ms between direct and reflected sounds is needed, which translates to a minimum distance of 1 m from the walls.

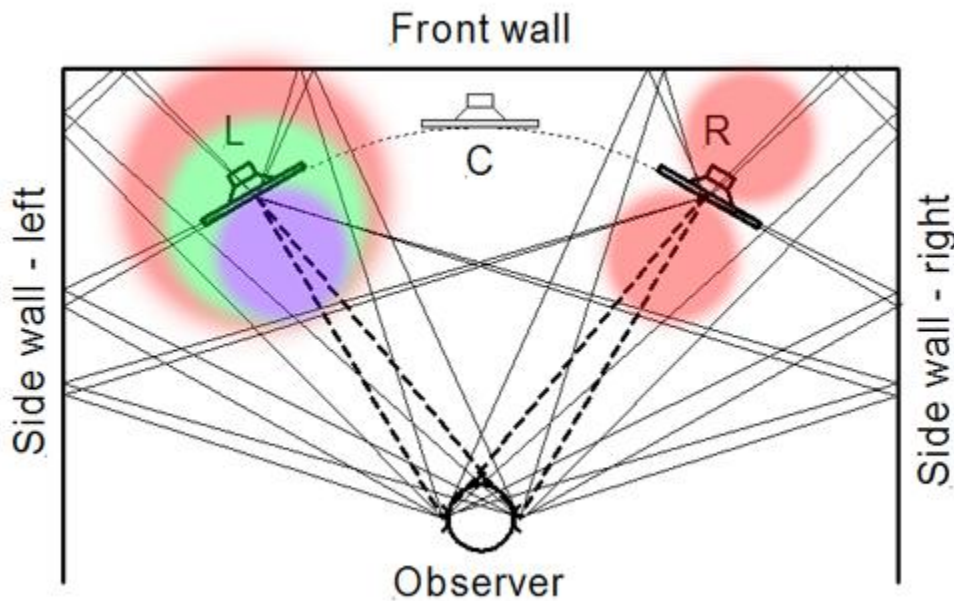


Fig. 26 Triangular loudspeaker and listener setup with symmetry relative to the walls and at some distance from them. The minimum distance is 1 m. The left loudspeaker will generate spectrally incoherent sound streams at the observer's ears.

Furthermore, reflected sounds should be spectrally coherent with the direct sound to a high degree for unambiguous perceptual processing. Direct and reflected sounds then generate a reverberant sound field that is spectrally coherent with the direct sound at the listener's ears [14, 15, 16]. Under those conditions a listener can withdraw attention from the room and focus on the aural scene in front of him. Room and loudspeakers disappear.

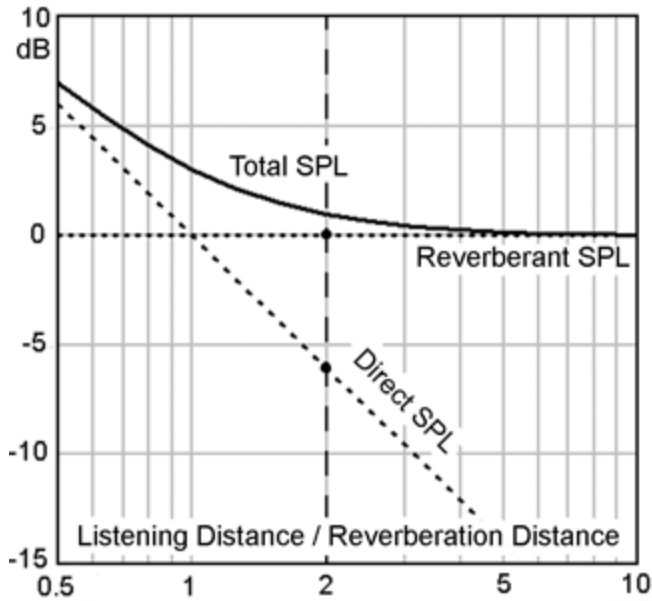


Fig. 27 SPL in a room as function of distance from the loudspeaker. The listening distance should be less than twice the reverberation radius for D/R less than -6 dB.

Coherence of the room sound requires a frequency independent polar response of the loudspeaker. It implies a constant directivity loudspeaker such as a monopole, dipole or cardioid. Experience has shown that the optimum listening distance should be less than twice the radiation radius so that the direct sound is less than 6 dB below the reverberant sound level in the room. In the example of Figure 23 that should be less than 1.8 m (6 ft) for the monopole or 3.1 m (10 ft) for the dipole source. The maximum recommended loudspeaker spacing would become 6 ft for the monopole or 10 ft for the dipole. The dipole reaches deeper into the room because it interacts less with it.

4.0 Summary

The type and performance of loudspeakers used, their setup, the listening distance and the room's acoustic properties determine the quality of the rendered aural scene at the consumer's end of the acoustical chain. The aural scene is ultimately limited by the recording.

The type and configuration of microphones used, their setup, the venue, the monitor loudspeakers for the mix, their setup, the listening distance, the studio's acoustic properties and the mix determine the quality of the recording at the producer's end of the acoustical chain between performer and consumer.

Monitor and consumer loudspeakers must exhibit close to constant directivity at all frequencies in order to obtain optimal results for recording and rendering.

5.0 References

[1] David Griesinger, **Frequency Response Adaptation in Binaural Hearing**, 126th AES Convention, Munich 2009, Preprint 7768, www.davidgriesinger.com

[2] Joerg Wuttke, **Zwei Jahre Kugelflaechenmikrofon**, Mikrofonbuch_Kap5.pdf, 1992, <http://www.schoeps.de/en/information>

- [3] Floyd E. Toole, **Sound Reproduction**, Focal Press, 2008
- [4] Peter Damaske, **Acoustics and Hearing**, Springer, 2008
- [5] Francis Rumsey, **Spatial Audio**, Focal Press, 2005
- [6] Helmut Wittek, **Image Assistant**, JAVA applet for determining the "Stereo Recording Angle", www.hauptmikrofon.de
- [7] SHOEPS Mikrofone, **Showroom**, www.schoeps.de/en/applications/showroom
- [8] Jens Blauert, **Spatial Hearing**, The MIT Press, 1997
- [9] Eric Benjamin, **An experimental Verification of Localization in Two-Channel Stereo**, 121st AES Convention, San Francisco 2006, Preprint 6968
- [10] Siegfried Linkwitz, **A Model for Rendering Stereo Signals in the ITD-Range of Hearing**, 133rd AES Convention, San Francisco 2012, Preprint xxxx
- [11] Juergen Meyer, **Acoustics and the Performance of Music**, Springer, 2009
- [12] Albert S. Bregman, **Auditory Scene Analysis - The Perceptual Organization of Sound**, The MIT Press, 1999
- [13] Heinrich Kuttruff, **Room Acoustics**, John Wiley & Sons, 1973
- [14] Siegfried Linkwitz, **Room Reflections Misunderstood?**, 123rd AES Convention, New York, October 2007, Preprint 7162, Abstract,
- [15] Brad Rakert, William M. Hartmann, "**Localization of sound in rooms. V. Binaural coherence and human sensitivity to interaural time differences in noise**", J. Acoust. Soc. Am., Vol. 128, No. 5, November 2010
- [16] William M. Hartmann, **Signals, Sound, and Sensation**, Springer, 2005
- [17] www.linkwitzlab.com, **Recording & Rendering**, 2012

2012, August 27
Siegfried Linkwitz

What you hear is not the air pressure variation in itself
but what has drawn your attention
in the two streams of superimposed air pressure variations at your eardrums